DOCUMENT RÈSUME

ED 197 317 CS 005 852

AUTHOR Harris, Albert J.: Jacobson, Milton D.
TITLE The Harris-Jacobson Readability Formula A.
PUB DATE May 80
NOTE 13p.: Paper presented at the Annual Meeting of the International Reading Association (25th, St. Louis, MO, May 5-9, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Education: *Measurement Techniques: *Readability: *Readability Formulas: *Reading Research
IDENTIFIERS *Harris Jacobson Readability Formula

ABSTRACT

This paper describes the Harris-Jacobson Readability Formula A, a revision of a readability formula based on word lists from "Basic Elementary Reading Vocabularies." Section one of the paper discusses procedures for using the formula, including the selection of samples, scoring the variables, and applying the formula. The next section discusses how the formula was developed and its characteristics. The final section reports research findings showing how Formula A compares to other readability formulas (Spache, Dale-Chall, and Flesch) and reveals results favorable to Formula A. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED197317

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Albert J. Harris
Milton D. Jacobson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Albert J. Harris
The City University of New York

and

Milton D. Jacobson
University of Virginia

The Harris-Jacobson Readability Formula A[1]

Special Interest Group on Readability, International Reading Association, St. Louis, May 6, 1980

Formula A is a revision of one of the two readability formulas developed by Harris and Jacobson (1974). Full directions for computing those formulas were presented in the 6th and 7th Editions of How to Increase Reading Ability (Harris and Sipay, 1975, 1980). Formula A is the latest of several formulas based on word lists from Basic Elementary Reading Vocabularies (Harris and Jacobson, 1972).

Formula A is a revision of the one called H-J Formula 2. The changes include: (1) replacing the H-J Short Readability Word List by the Long Readability Word List;[2] (2) counting an uncommon word each time it appears in a sample, rather than just the first time; and (3) adding a Table Conversion Value that improves the fit of

---

[1]Formula A was completed before the untimely death of Dr. Jacobson in September, 1980. The present paper is a revision of the one that he presented on May 6, 1980.

[2]The Short List contained all first- and second-grade words in the Core and Additional lists, 912 root words and 1800 inflected forms, totalling 2,792 items. The Long List included third-grade words in addition to the first- and second-grade words, 1,925 root words and 4,076 inflected forms, totalling 6,001 items.

CS005852

the formula, particularly above primary reading levels. The effect of these changes is to produce a formula that works well over a very wide range, and is quick and easy to use.

Procedure

Selecting Samples

Since the readability of a single selection or entire book is estimated from measured characteristics of a small number of samples, it is important that the samples chosen be representative of the entire selection or book. Even for a single selection, a minimum of three samples should be taken and the results averaged. For a book, at least five samples should be taken. The samples should be spread at approximately equal intervals through the material. For three samples, choose one sample from the first third, one from the middle third, and one from the final third. For five samples, choose one from each fifth of the work. Having located the part of the material to be sampled, open the material at random to a page within that portion. Do not use the first two paragraphs of a selection in which there is often a concentration of new vocabulary, and do not count verbal material in figures or tables.

Start each sample with the beginning of a paragraph, marking the beginning with a check in the margin. Count 200 consecutive words and continue counting to the end of the sentence that contains the 200th word. Do not count numbers or other nonverbal symbols. If one is in a hurry samples of about 100 words can be used, but the result will be less reliable.

Scoring the Variables

Variable 1. Variable 1 (V1) is the percent of unfamiliar words in the sample. For this formula, an unfamiliar word is defined as a word that is not listed as third grade level or below in Basic Elementary Reading Vocabularies (A.J. Harris and Jacobson, 1972). The total Alphabetical List which begins on page 107 of that book includes 6,866 general words, plus some technical words, in one alphabetical sequence. Inflected forms that are listed with their root words are of the same level as the root word. Words listed as third grade level or below are considered familiar; words listed as fourth grade or higher, words listed without a grade designation, and words missing from the list are counted as unfamiliar. Proper nouns are not counted as unfamiliar.

Some examples may clarify these points. Act, acted, acting, and acts are (C2), second grade level, and are familiar. Admire, admired, admiring, and admires are (C4), fourth grade level, and are unfamiliar. Acceptable (EN) does not have a grade designation and is therefore unfamiliar. Anticipate is missing from the list and counted as unfamiliar. An unfamiliar word that appears more than once in a sample is counted each time it appears.

Start at the beginning of the sample and look up each possibly unfamiliar word in the Total Alphabetical List. If it is unfamiliar make a tally mark (/). Do the same for each possibly unfamiliar word. Count the number of tallies and enter the result on the Worksheet.

When first learning to use Formula A one will check the Total Alphabetical List for many words. With practice, an increasing number of words can be identified as familiar or unfamiliar without looking them up in the list, and scanning the list for a particular word becomes quite fast.

To find the numerical value of V1, divide the number of tallies by the number of words in the sample and multiply by 100 to express as a percent. The use of a small calculator speeds up this and other calculations in using Formula A.

Variable 2. This variable (V2) is average sentence length, more precisely described as the mean number of words per sentence. Count the number of sentences in the sample. Headings or subheadings are counted as sentences. The V2 score is obtained by dividing the number of words in the sample by the number of sentences, carrying the result to three decimal places.

Applying the Formula

Formula A in regression equation form is as follows:

Estimated Readability Level = V1 x 0.14 + V2 x 0.153 + 0.56 + Table Conversion Value

The Table Conversion Value may be found in Table 1. Compute Formula A through the addition of 0.56; this sum is called the Raw Score. Find the Raw Score interval that contains the obtained Raw Score; the corresponding Table Conversion Value is shown in the table.

Table 1. Table Conversion Values for Formula A Corresponding to Raw Scores

Raw Score	Table Conversion Value
less than 1.99	0.0
2.0 - 2.49	0.2
2.5 - 2.99	0.4
3.0 - 3.49	1.2
3.5 - 3.99	2.0
4.0 - 4.49	2.3
4.5 - 4.99	2.4
5.0 - 5.99	3.0
6.0 - 6.99	3.5
7.0 - 7.99	4.0
8.0 - 8.99	4.5

Some examples of the computation of Formula A are shown in the Worksheet. In ordinary use, the five columns of the Worksheet would be used for five samples from the same book. This Worksheet has one sample from each of five basal readers in the same series.

Worksheet for H-J Readability Formula A

Title of Book:

	Samples				
	1[2]	3[1]	4	5	6
Starting on page	123	60	290	213	193
(1) A. No. of words	202	214	207	201	200
(1) B. No. of unfamiliar words	1	10	11	15	26
(1) C. No. of sentences	23	24	15	16	12
(2) V1 = B ÷ A x 100	0.5	4.673	5.314	7.463	13.0
(3) V2 = A ÷ C	8.783	8.914	13.80	12.563	16.667
(4) V1 x 0.14	0.069	0.654	0.744	1.045	1.820
(5) V2 x 0.153	1.344	1.364	2.111	1.922	2.550
(6) Add constant	0.560	0.560	0.560	0.560	0.560
(7) Raw Score	1.973	2.578	3.415	3.527	4.930
(8) Add Table Conv. Val.	0.0	0.4	1.2	2.0	2.4
Estimated Readability Level*	2.0	3.0	4.6	5.5	7.3

*It is desirable to carry three decimal places up to the final step and then round to one decimal place.

The steps in computation are as follows: (1) Enter the number of words, number of unfamiliar words, and number of sentences in the sample. ~~(2) Obtain V1 by dividing the number of sentences in the sample.~~ (2) Obtain V1 by dividing the number of unfamiliar

words by the total number of words and multiplying by 100. (3) Obtain V2 by dividing total number of words by number of sentences. (4) Multiply V1 by 0.14. (5) Multiply V2 by 0.153. (6) Get the sum of the result of step (4), the result of step (5), and 0.56; this is the Raw Score. (7) Enter the Table Conversion Value corresponding to the Raw Score. (8) Add the Table Conversion Value to the Raw Score; this gives the Estimated Readability Level.

It may be noted that a sample from a very easy source may contain no unfamiliar words. In these cases the value of V1 is zero and the result of step (4, is zero.

Formula Development and Characteristics

Following the publication of Basic Elementary Reading Vocabularies (A.J. Harris and Jacobson, 1972), five Harris-Jacobson readability formulas were developed (A.J. Harris and Jacobson, 1974). Two of those formulas were published, with full directions for computation, in the Sixth and Seventh Editions of How to Increase Reading Ability (1975, 1980). These were both two-variable formulas that used a measure of unfamiliar words and average sentence length.

The measure of unfamiliar words was the percent of words not found in the H-J Short List (first- and second-grade words); its correlation with the criterion of average basal reader level was .80 for grades 1-3 and .87 for grades 1-6. In the 1974 report it was noted that the Short List was a bit better than the Long List (first-, second-, and third-grade words) at primary grades, but the two lists were equally good at middle grade levels and the Long List would probably be better above sixth grade level.

The other variable was mean number of words per sentence. This correlated .83 with the criterion for grades 1-3 and .79 for grades 1-6. The two variables were given weights to produce the maximum possible multiple correlation, one set of weights for grades 1-3 and a somewhat different set of weights for grades 1-6.

Formula 1, recommended for use with primary-level material, correlated .90 with the criterion for grades 1-3, and Formula 2, recommended for use with middle grade materials, correlated .90 for grades 1-6. Standard errors of estimate were .38 for Formula 1 and .71 for Formula 2. The two formulas did not always agree on the ratings given materials intended for third and fourth grade use.

Several other Harris-Jacobson readability formulas were developed using McCall, Crabbs comprehension scores as the criterion (A.J. Harris and Jacobson, 1975, 1976). These were multi-variable formulas that required use of special computer programs devised by Dr. Jacobson, and included a number of spelling patterns. The Long List showed higher correlation with comprehension scores at grades 4, 6 and 9, and was used in those formulas.

In developing Formula A the previous Formula 2 was modified to take account of two problems: the difficulty at third and fourth grade levels, and the superiority of the Long List above third grade. The Long List was tried in two ways, with unique words (an unfamiliar word was counted only the first time it appeared in a sample) and with total words (the word was counted as many times as it appeared). The total word count turned out

to be slightly the better. The other change was to develop a Table Conversion Value so that at fourth grade level the formula would provide an average of the grade equivalents of primary readability formulas (Harris-Jacobson Formula 1 and Spache 1953) and the intermediate formula of Dale and Chall. The adjustment values were determined by processing the modified formula through one entire reading series, preprimer through grade eight.

Comparison of Formula A with Other Readability Formulas

Two pre-publication sets of basal readers from preprimer through eighth grade (Economy 1980, Houghton Mifflin 1981) were combined to provide a criterion based on a total of 1,311,337 running words of text, divided into samples of 200-plus words. Readability scores for the average of both texts at each level were computed for Formula A, two primary-grade readability formulas (H-J Formula 1 and Spache 1953), and two formulas used mainly for materials from fourth grade up (Dale-Chall and Flesch). The results are shown in Table 2. The assumption is made that the closer a formula comes to the average level indicated by the publishers, the better.

Formula A and H-J Formula 1 show extremely similar results at first and second grade levels, with Formula A fitting better at third grade levels.

Formula A has a better fit than the Spache formula at first and high third levels, with quite similar results for the two formulas at second and low third levels.

Table 2. Mean Readability Scores for Formula A and Four Other Readability Formulas

Reader Level	H-J Formula 1	Spache '53 Formula	H-J Formula A	Dale-Chall	Flesch
pp 1	1.2	1.5	1.3		2.3
pp 2	1.3	1.8	1.4		2.5
pp 3	1.4	1.8	1.5		2.6
1^1	1.7	2.0	1.7		4.2
1^2	2.0	2.2	2.0		5.1
2^1	2.3	2.5	2.4		5.7
2^2	2.7	2.8	2.8		6.0
3^1	3.0	3.2	3.2		6.3
3^2	3.3	3.5	3.9		6.5
4			4.4	5.2	6.5
5			5.8	5.8	7.0
6			6.5	6.3	7.1
7			7.7	6.9	7.4
8			8.5	7.1	7.6

Formula A and the Dale-Chall formula agree very closely at fifth and sixth grade levels. The Dale-Chall formula overestimated the difficulty of the fourth readers and underestimated the seventh and eighth readers; that is, however, before application of the Dale-Chall adjustments.

The Flesch formula, which has been used mainly with adult materials, severely overestimated the difficulty of the basal readers from first through fifth grade levels, by as much as 3.3 grades. The Flesch formula fitted well at sixth, seventh, and eighth grade levels, but not as closely as Formula A.

These comparisons show that Formula A compares favorably with two widely-used primary grade readability formulas and two widely-used formulas for the middle and upper grades. It provides a continuous scale from beginning first grade through eighth grade, based on the measured characteristics of the criterion texts. Estimated Readability Levels from ninth grade up through grade thirteen can be obtained, but these are based on extrapolation. Formula A provides a significant improvement over the previously published H-J Formulas 1 and 2.

REFERENCES

Harris, Albert J., and Ja obson, Milton D. Basic Elementary Reading Vocabularies. New York: Macmillan, 1972.

Harris, Albert J., and Jacobson, Milton D. Revised Harris-Jacobson Readability Formulas. Paper presented at the annual convention of the College Reading Association, Bethesda, Maryland, October 31, 1974.

Harris, Albert J., and Jacobson, Milton D. Comparative Predictions of Tested Reading Comprehension of Intermediate and Secondary Pupils. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., April 1, 1975.

Harris, Albert J., and Jacobson, Milton D. Predicting Twelfth Graders' Comprehension Scores. Journal of Reading, October 1976, 20, 43-46.

Harris, Albert J., and Jacobson, Milton D. A Framework for Readability Research: Moving Beyond Herbert Spencer. Journal of Reading, February 1979, 22, 390-398.

Harris, Albert J., and Jacobson, Milton D. A Comparison of the Fry, Spache, and Harris-Jacobson Readability Formulas for the Primary Grades. The Reading Teacher, May 1980, 33, 920-923.

Harris, Albert J., and Sipay, Edward R How to Increase Reading Ability, 6th and 7th Editions. New York: Longman Inc., 1975, 1980.